


**Science  
MAGAZINE**

HOME

HELP

SEARCH

ARCHIVE

SUBSCRIPTIONS

FEEDBACK

ORDER an article

SIGN IN

SUSAN T. ARNOLD || [Change Password](#) || [View/Change User Information](#) || [CiteTrack Personal Alerts](#) ||  
[Subscription HELP](#) || [Sign Out](#)

## Stephen P. A. Fodor

### Techwire

Using the URL at the end of this item, readers can immediately offer feedback and suggestions on this topic.

- Search Medline for articles by:  
Fodor, S. P.
- Alert me when:  
New articles cite this article

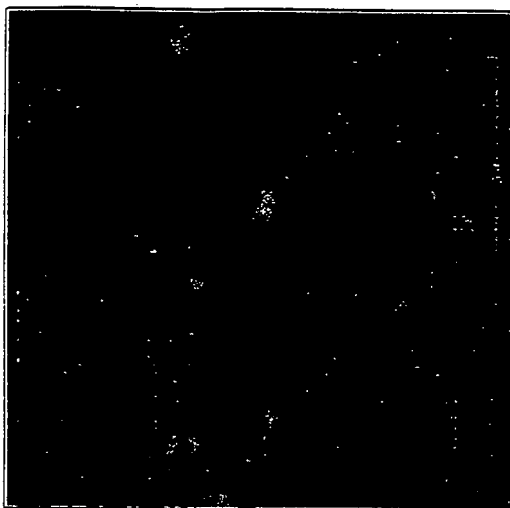
An enormous amount of genetic sequence data has been obtained with gel-based DNA sequencing methods [HN1]. To understand the function of the genes and their health implications, genetic variation over vast numbers of cells, tissues, individuals, and organisms must be examined. The complex expression patterns of the estimated 100,000 genes comprising the human genome [HN2] and the intricate developmental signals that define cellular fate will need to be understood. The magnitude of this analysis will dwarf the task of obtaining the primary sequence of the human genome, and so efficient means to experimentally access vast amounts of genetic information are critically needed.

Conventionally, researchers use analytical techniques to resolve sequence at the single-nucleotide level (1). In contrast, biological systems read, store, and modify genetic information, using the simple rules of molecular recognition. Each DNA strand carries the capacity to recognize uniquely complementary sequence through base pairing. The process of recognition, or hybridization [HN3], is highly parallel, and every sequence in a complex mixture can, in principle, be interrogated at the same time. Application of this highly desirable concept to sequence analysis has awaited new combinatorial technologies to generate high-density ordered arrays of large numbers of oligonucleotide probes (2).

At Affymetrix [HN4], we developed ways to synthesize and assay biological molecules in a highly dense parallel format. Integration of two key technologies forms the cornerstone of the method (3). The first technology, light-directed combinatorial chemistry [HN5], enables the synthesis of hundreds of thousands of discrete compounds at high resolution in precise locations on a substrate. The second, laser confocal fluorescence scanning [HN6], permits measurement of molecular interactions on the array. This technology is now commercial, with complete systems exported to dozens of sites.

Light-directed chemical synthesis employs two mature technologies: photolithography [HN7] and solid-phase synthesis [HN8]. Synthetic linkers modified with photochemically removable protecting groups are attached to a glass substrate. Light is directed through a photolithographic mask to specific areas of the surface to produce localized photodeprotection. The first of a series of chemical building blocks--hydroxyl-protected deoxynucleosides, for example--is incubated with the surface, and chemical coupling occurs at those sites that have been illuminated in the preceding step. Next, light is directed to a different region of the substrate by a new mask, and the chemical cycle is repeated. Highly efficient strategies can be used to synthesize any arbitrary probe at any discrete, specified location on the array in a minimum number of chemical steps. For example, the complete set of  $4^N$  polydeoxynucleotides of

length  $N$ , or any subset of this set, can be synthesized in only  $4N$  chemical cycles. Thus, given a reference sequence, a DNA chip can be designed that consists of a highly dense array of complementary probes with no restriction on design parameters. The amount of nucleic acid information encoded on the chip in the form of different probes is limited only by the physical size of the array and the achievable lithographic resolution. Current bulk manufacturing methods allow for ~409,000 polydeoxynucleotides to be synthesized on 1.28-cm by 1.28-cm chips.



**Arrayed for sequencing.** DNA chip fabricated by photolithography.

Photolithography allows the construction of probe arrays with extremely high information content. Because the array is constructed on glass, it can be inverted and mounted in a temperature-controlled hybridization chamber. A target sequence is fluorescently tagged and then injected into the chamber, where the target hybridizes to its complementary sequences on the array. Laser excitation [HN9] enters through the back of the array, focused at the interface of the array surface and the target solution. Fluorescence emission is collected by a lens and passes through a series of optical filters to a sensitive detector. By simply scanning the laser beam or translating the array, or a combination of both, a quantitative two-dimensional fluorescence image of hybridization intensity is rapidly obtained. Commercial instrumentation for controlling the hybridization and scanning of the arrays, and software for image and data analysis have been developed. This approach requires only minute consumption of chemical reagents and minute preparations of biological samples.

An array of oligonucleotides complementary to subsequences of a target sequence can be used to identify a target sequence, measure its amount or relative expression level, and detect differences between the target and a reference sequence. Many different arrays can be designed for these purposes, and the applications appear to be only limited by imagination. The system consists of chips, a hybridization station to control hybridization, and a reader and software to access the chip data. Specific chip products for expression analysis, HIV array resistance screening, and gene resequencing are already on the market. Two versions of commercial readers are available: a first-generation system from Molecular Dynamics [HN10] as well as a recently released high-performance system from Hewlett-Packard [HN11]. Chip production is now in a scalable format. We are now producing ~5,000 to 10,000 chips per month, and we expect a large increase in production in the near future.

To fully understand gene expression, gene function, and the subtleties of regulation, the quantitative levels of expressed genes under various conditions must be assayed. In addition, if quantitative "snapshots" of gene expression can be captured, the dynamics of cellular pathways can then be deciphered. Recently, Lockhart *et al.* (4), published methods for the quantitative parallel measurement of cellular messenger RNA for gene sequences encoded on the chip solely from primary sequence data. RNAs present at a frequency of 1:300,000 were unambiguously detected with a quantitative assay spanning three to four orders of magnitude in concentration. Currently, Lockhart and group have developed chips containing the complete open reading frames from the yeast genome [HN12], a series of "custom" chips with hundreds to thousands of full-length genes or fragments from various databases, as well as "standard" chips containing more than 6,500 genes from the public databases. An expression chip with more than 50,000 expressed sequence tags from the public databases [HN13] is currently in development.

This method complements other noncombinatorial array-based methods that involve the serial spotting of multiple clones or complementary DNAs (cDNAs) onto nylon membranes or modified microscope slides (5). These latter methods are inherently parallel for the analysis of sequence information, and they complement the probe-based arrays in their ability to use previously nonsequenced biological materials for the expression analysis. However, hybridization to large cDNA or PCR products can be thermodynamically more stable than to a series of shorter discriminating oligonucleotides and can yield confusing cross-hybridization signals when examining closely related genes, gene families, or other variants. Nonetheless, as experimentation continues, new ways around these problems will be found. In combination, the synthetic and mechanical techniques build a powerful new platform for parallel gene expression and offer large time and labor improvements over traditional cDNA library sequencing methods [HN14].

Understanding the relationship between genotype and phenotype is a critical technical bottleneck in modern genetics. For example, consider examining 50 kilobases (kb) of coding sequence for 1000 individuals. The genes are known, but the prevalence, location, and identity of polymorphisms are not. The methods of conventional gel-based sequencing that are so effective in the initial gene sequencing are not efficient for this task. Comparative gel-based sequencing is indistinguishable from a de novo sequencing reaction, and so the de novo sequencing reaction must be carried out for all 50 kb over the 1000 individuals, or roughly 50 Mb of sequence. This illustrates the stark contrast of technical requirements when compared to a chip-based approach.

Chee *et al.* (6) recently showed how the entire human mitochondrial genome [HN15] can be sequenced with high accuracy in a single hybridization experiment. A total of ~135,000 oligonucleotide probes were used to check the sequence of ~33 kb (forward and reverse strands) of the mitochondrial genome in one reaction. In addition, 179 of the 180 polymorphisms present in control samples were correctly detected. Two-color comparative sequence analysis experiments were performed that demonstrated how mutations or polymorphisms could be detected on a very large scale, making it now possible to use the technology for large-scale polymorphism screening efforts. At the current state of development, 1.28-cm by 1.28-cm chips can contain enough probes to scan anywhere from 32 kb to more than several hundred kilobases of sequence, depending on the specific chip design and accuracy requirements of the screen. Put in the context of the previously posed experiment, 1000 chips each containing 50 kb could easily and quickly perform the comparative sequence analysis.



**Making the matrix.** Glass wafer is divided into DNA chips that contain the probe arrays.

Designing arrays to detect specific allelic variation is relatively straightforward. In addition to using chip designs appropriate to scan a sequence (as in the polymorphism application), blocks of probes can be

dedicated to the specific detection of known allelic variation. Cronin *et al.* (7) designed a chip-based assay to detect multiple mutations in the CFTR gene, Kozal *et al.* (8) targeted HIV [HN16], Hacia *et al.* (9) examined the BRCA1 gene [HN17], and a number of new designs are in development for examination of p53 and cytochrome p450, and for microbial identification and antibiotic resistance. The amount of data coded on the array is limited only by the number of probes used per data point, the available synthesis area, and the synthesis resolution.

Recently, in collaboration with Lander's group at the Whitehead Institute (Cambridge, MA) [HN18], Chee and Lipshutz have initiated work on a single nucleotide polymorphism (SNP) mapping chip (10). The immediate objective is to identify the common polymorphisms (those of ~20 to 50% frequency) contained within the mapped sequence-tagged site collection at the Whitehead Institute. These then form the basis set of biallelic markers that can be amplified from genomic DNA and applied to a probe array. Similar to the design used to detect allelic variation in the CFTR gene, blocks of probes are dedicated to each polymorphic form of the marker. This allows a straightforward detection of whether the sample is homozygous or heterozygous for each marker. These experiments offer enormous savings in time and labor, compared to standard gel-based microsatellite methods. Currently, prototype mapping chips containing ~500 markers are being produced, with plans to expand to a 2000-marker chip by the end of the year. These chips will be used for a number of applications, including studies of linkage, association, and loss of heterozygosity measurements.

The challenges of linking sequence variation to biological function are many. When the sequence is available, chips containing every gene in the human genome (chips for other genomes such as yeast have already been manufactured) can be produced, allowing genome-wide expression analysis. This should have a profound influence on the ability to elucidate the metabolic and disease pathways [HN19] of the cell under a variety of developmental and environmental perturbations and have immediate applications in toxicology studies and pharmaceutical development [HN20]. Screening chips will allow the databasing of large numbers of polymorphisms, and SNP chips will uncover how they are associated with disease. Chips providing insight into the genetics of model organisms have already been developed with strategies that will no doubt expand to include any organism of interest (11). DNA chip technology moves genetic sequence analysis away from serial gel-based methods to a massively parallel screening format. In time, technology will be needed to make this same paradigm shift for the hundreds of thousands of proteins, chemical messengers, and other molecular components of life.

TechWire Forum: [www.sciencemag.org/dmail.cgi?53241](http://www.sciencemag.org/dmail.cgi?53241)

## References and Notes

1. F. Sanger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463 (1977) [Medline]; A. M. Maxam *et al.*, *ibid.*, p. 560. [Medline]
2. S. P. A. Fodor *et al.*, *Science* 251, 767 (1991) [Medline]; A. C. Pease *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 91, 5022 (1994). [Medline] Techniques to generate and use other oligonucleotide arrays have also been described but are not covered here; see E. Southern, U. Maskos, and R. Elder [Genomics 13, 1008 (1992) [Medline]] and E. Southern [Trends Genet. 12, 110 (1996) [Medline]].
3. S. P. A. Fodor *et al.*, *Nature* 364, 555 (1993). [Medline]
4. D. L. Lockhart *et al.*, *Nature Biotechnol.* 14, 1675 (1996).
5. S. Meier-Ewert *et al.*, *Nature* 361, 375 (1993) [Medline]; C. Nguyen *et al.*, *Genomics* 29, 207 (1995) [Medline]; N. Takahashi *et al.*, *Gene* 164, 219 (1995) [Medline]; A. Milosavljevic *et al.*, *Genome Res.* 6, 132 (1996) [Medline]; M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *Science* 270, 467 (1995). [Medline]
6. M. Chee *et al.*, *Science* 274, 610 (1996). [Medline]
7. M. T. Cronin *et al.*, *Hum. Mutat.* 7, 244 (1996). [Medline]
8. M. Kozal *et al.*, *Nature Med.* 2, 753 (1996). [Medline]
9. J. Hacia *et al.*, *Nature Genet.* 14, 441 (1996). [Medline]
10. E. S. Lander, *Science* 274, 536 (1996). [Medline]
11. D. Shoemaker, *Nature Genet.* 14, 450 (1996). [Medline]

12. Supported in part by NIH grant P01HG01323.

---

*The author is with Affymetrix, Santa Clara, CA 95061, USA. E-mail: [steve\\_fodor@affymetrix.com](mailto:steve_fodor@affymetrix.com)*

---

## HyperNotes

Related Resources on the World Wide Web

### Numbered Hypernotes

1. This site, located on the Web server of George Church's laboratory at Harvard University, presents a convenient jumping point to all of the Web genome sequencing projects. Links to most of the organisms that are currently being sequenced can be found here—from worms to humans.
2. The Genome Database is a major site on the Web where different types of data collected on the human genome are being archived. This site collects and organizes regions of the human genome, clones, amplicons (PCR markers), breakpoints, cytogenetic markers, fragile sites, expressed sequence tags (ESTs), syndromic regions, and contigs. It also houses current maps of the human genome, including cytogenetic maps, linkage maps, radiation hybrid maps, content contig maps, and integrated maps.
3. This primer on molecular genetics has been created by the U.S. Department of Energy. It gives excellent background data on the molecular basis of DNA, genes, and chromosomes and is suited for both beginners and experts.
4. The home page of Affymetrix gives a listing of the current GeneChip products used for chip-based DNA detection as well as background information on the research activities of the company.
5. Interested in learning more about chemistry or participating in online discussions with other chemists? The home page of the Royal Society of Chemistry provides a wealth of information, including a list of chemistry servers on the Internet.
6. To learn more about confocal microscopy techniques, see this site presented by L. Ladic from the University of British Columbia. There is a basic introduction of the techniques and tips for specimen preparation.
7. Photolithography is used in industry to produce the integrated circuits that serve as the brains of computers. This tutorial provides information on how the process works and on how it is evolving.
8. This site **collects** references to the most recent papers published in the field of solid-phase synthesis **and** combinatorial chemistry. It also has a large collection of links to companies that sell products for solid-phase synthesis.
9. The Beckman Laser Institute and Medical Clinic at the University of California at Irvine has collected a large amount of basic information on the application of lasers to research and medicine. Examples include how lasers are used to remove wrinkles or improve vision.
10. The home page of Molecular Dynamics provides information on advanced imaging techniques for biology and chemistry. The online discussion is a good way to get the latest tricks and tips on using the imaging techniques.
11. The Hewlett-Packard Web site has some interesting surprises, including a focus on the basic science of chemistry. Their Internet chemistry resource list is useful to basic scientists as well as commercial researchers.

12. To learn about almost every database available on the Internet for the yeast *Saccharomyces cerevisiae*, turn to this site at the National Institutes of Health. Here you can get any portion of the completely sequenced genome as well as submit queries to the XREF database.
13. The National Center for Biotechnology Information at the National Institutes of Health is the gateway to the world's public databases housing nucleic acid and protein sequences.
14. Traditional high-throughput sequencing relies heavily on the use of automated sequencing equipment. The Applied Biosystems machine is currently one of the most frequently used. This Web site describes some of its features.
15. MITOMAP, provided by A. M. Kogelnik, M. T. Lott, M. D. Brown, S. B. Navathe, and D.C. Wallace of Emory University and the Georgia Institute of Technology (Atlanta, GA), serves as a home base for studies on human mitochondrial DNA(mtDNA). It houses the entire mtDNA sequence, mutation collections, and population data.
16. I. Fenton, of the University of Wales College of Medicine (UK), maintains the Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff, which collects and maintains data on many of the known mutations in human genes that cause diseases, including cystic fibrosis.
17. The Breast Cancer Information Core houses a mutation database of known mutations in the gene BRCA1. Although the site can be accessed by members only, membership is free and open to the scientific public by following the guidelines presented and filling out their form.
18. The home page of the Whitehead Institute for Biomedical Research provides this link to their genome center, a joint project with the Massachusetts Institute of Technology.
19. V. A. McKusick and his colleagues at Johns Hopkins University and elsewhere provide the Online Mendelian Inheritance of Man site, which collects a wealth of information about individual genetic defects that cause human diseases. The short summary articles provide a good way to quickly find out about the current knowledge of a specific genetic defect.
20. One way to find out what the different pharmaceutical companies are up to is by checking out their Web sites. Links to many of them are available at Pharm Web.

► Search Medline for articles by:  
Fodor, S. P.  
► Alert me when:  
New articles cite this article

Volume 277, Number 5324 Issue of 18 Jul 1997, pp. 393 - 395  
©1998 by The American Association for the Advancement of Science.

|             |          |                  |             |         |         |
|-------------|----------|------------------|-------------|---------|---------|
| SCIENCE NOW | HOME     | HELP             | SEARCH      | ARCHIVE | CAREERS |
| NEXT WAVE   | FEEDBACK | ORDER an article | MARKETPLACE |         |         |

Copyright © 1998 by the American Association for the Advancement of Science.

**PNAS**  
IS NOW ONLINE!

Clontech  
Plasmid  
Purification

We've Surf'd  
For You

**Science**  
MAGAZINE

HOME

HELP

SEARCH

ARCHIVE

SUBSCRIPTIONS

FEEDBACK

ORDER an article

SIGN IN

SUSAN T. ARNOLD || [Change Password](#) || [View/Change User Information](#) || [CiteTrack Personal Alerts](#) ||  
[Subscription HELP](#) || [Sign Out](#)

## Accessing Genetic Information with High-Density DNA Arrays

Mark Chee, Robert Yang, Earl Hubbell, Anthony Berno, Xiaohua C. Huang, David Stern, Jim Winkler, David J. Lockhart, Macdonald S. Morris, Stephen P. A. Fodor

- ▶ [PubMed Citation](#)
- ▶ [Related Articles in PubMed](#)
- ▶ [Download to Citation Manager](#)
- ▶ This article has been cited by [other articles](#)
- ▶ Search Medline for articles by:  
    [Chee, M.](#) || [Fodor, S. P. A.](#)
- ▶ Alert me when:  
    [New articles cite this article](#)

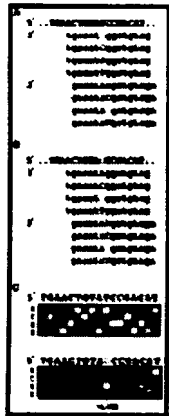
Rapid access to genetic information is central to the revolution taking place in molecular genetics. The simultaneous analysis of the entire human mitochondrial genome is described here. DNA arrays containing up to 135,000 probes complementary to the 16.6-kilobase human mitochondrial genome were generated by light-directed chemical synthesis. A two-color labeling scheme was developed that allows simultaneous comparison of a polymorphic target to a reference DNA or RNA. Complete hybridization patterns were revealed in a matter of minutes. Sequence polymorphisms were detected with single-base resolution and unprecedented efficiency. The methods described are generic and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability.

Affymetrix, 3380 Central Expressway, Santa Clara, CA 95051, USA.

A central theme in modern genetics is the relation between genetic variability and phenotype. To understand genetic variation and its consequences on biological function, **an enormous effort in comparative sequence analysis will need to be carried out.** Conventional nucleic acid sequencing technologies make use of analytical separation techniques to resolve sequence at the single nucleotide level (1, 2). However, the effort required increases linearly with the amount of sequence. In contrast, biological systems read, store, and modify genetic information by molecular recognition (3). Because each DNA strand carries with it the capacity to recognize a uniquely complementary sequence through base pairing, the process of recognition, or hybridization, is highly parallel, as every nucleotide in a large sequence can in principle be queried at the same time. Thus, hybridization can be used to efficiently analyze large amounts of nucleotide sequence. In one proposal, sequences are analyzed by hybridization to a set of oligonucleotides representing all possible subsequences (4). A second approach, used here, is hybridization to **an array** of oligonucleotide probes designed to match specific sequences. In this way the most informative **subset** of probes is used. Implementation of these concepts relies on recently developed combinatorial technologies to generate any ordered array of a large number of oligonucleotide probes (5).

The fundamentals of light-directed oligonucleotide array synthesis have been described (5, 6). Any probe can be synthesized at any discrete, specified location in the array, and any set of probes composed of the four nucleotides can be synthesized in a maximum of  $4N$  cycles, where  $N$  is the length of the longest probe in the array. For example, the entire set of  $\sim 10^{12}$  20-nucleotide oligomer probes, or any desired subset, can be synthesized in only 80 coupling cycles. The number of different probes that can be synthesized is limited only by the physical size of the array and the achievable lithographic resolution (7).

An array consisting of oligonucleotides complementary to subsequences of a target sequence can be used to determine the identity of a target sequence, measure its amount, and detect differences between the target and a reference sequence. Many different arrays can be designed for these purposes. One such design, termed a 4L tiled array, is depicted in Fig. 1A. In each set of four probes, the perfect complement will hybridize more strongly than mismatched probes. By this approach, a nucleic acid target of length  $L$  can be scanned for mutations with a tiled array containing  $4L$  probes. For example, to query the 16,569 base pairs (bp) of human mitochondrial DNA (mtDNA), only 66,276 probes of the possible  $\sim 10^9$  15-nucleotide oligomers need to be used.



**Fig. 1. (A)** Design of a 4L tiled array. Each position in the target sequence (uppercase letters) is queried by a set of four probes on the chip (lowercase letters), identical except at a single position, termed the substitution position, which is either A, C, G, or T (blue indicates complementarity, red a mismatch). Two sets of probes are shown, querying adjacent positions in the target. **(B)** Effect of a change in the target sequence. The probes are the same as in (A), but the target now contains a single-base substitution (base C, shown in green). The probe set querying the changed base still has a perfect match (the G probe). However, probes in adjacent sets that overlap the altered target position now have either one or two mismatches (red) instead of zero or one, because they were designed to match the target shown in (A). **(C)** Hybridization to a 4L tiled array and detection of a base change in the target. The array shown was designed to the mt1 sequence. (Top) hybridization to mt1. The substitution used in each row of probes is indicated to the left of the image. The target sequence can be read 5' to 3' from left to right as the complement of the substitution base with the brightest signal. With hybridization to mt2 (bottom), which differs from mt1 in this region by a T→C transition, the G probe at position 16,493 is now a perfect match, with the other three probes having single-base mismatches (**A** 5, **C** 3, **G** 37, **T** 4 counts). However, at flanking positions, the probes have either single- or double-base mismatches, because the mt2 transition now occurs away from the query position.

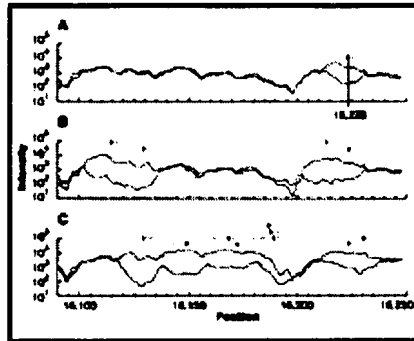
[View Larger Version of this Image (30K GIF file)]

The use of a tiled array of probes to read a target sequence is illustrated in Fig. 1C. A tiled array of 15-nucleotide oligomers varied at position 7 from the 3' end ( $P^{15,7}$ ) was designed and synthesized for mt1, a cloned sequence containing 1311 bp spanning the control region of mtDNA (8, 9, 10, 11). The upper panel of Fig. 1C shows a portion of the fluorescence image of an array hybridized with fluorescein-labeled mt1 RNA (12). The base sequence can be read by comparing the intensities of the four probes within each column. For example, the column for position 16,493 consists of the four probes, 3'-TGACAT**A**GGCTGTAG, 3'-TGACAT**C**GGCTGTAG, 3'-TGACAT**G**GGCTGTAG, and 3'-TGA- CAT**T**GGCTGTAG. The probe with the strongest signal is the probe with the **A** substitution (**A**, 49 counts; **C**, 8 counts, **G**, 15 counts, and **T**, 8 counts, where the background is 2 counts), identifying the base at position 16,493 as U in the RNA transcript. Continuing the process, the sequence at each position can be read directly from the hybridization intensities.

The effect on the array hybridization pattern caused by a single base change in the target is illustrated in Fig. 1B, and the detection of a single-base polymorphism is shown in the lower panel of Fig. 1C. The target was mt2, which differs from mt1 in this region by a T-to-C transition at position 16,493. Accordingly, the probe with the **G** substitution (third row) displays the strongest signal. Because the tiled array was designed to complement mt1, the hybridization intensities of neighboring probes that overlap position 16,493 are also affected by the change in target sequence. The hybridization signals of 15 probe sets of the 15-nucleotide oligomer tiled array are perturbed by a single base change in the target sequence. In the  $P^{15,7}$  array, each probe querying the eight positions to the left and six positions to the right of the polymorphism contain at least one mismatch to the target. The result is a characteristic loss of signal or a "footprint" for the probes flanking a mutation position. Of the four probes querying each position, the loss of signal is greatest for the one designed to match mt1. We denote the subset of probes with zero mismatches to the reference sequence as  $P^0$ .

A comparison of  $P^0$  hybridization signals from a target to those from a reference is ideally obtained by hybridizing both samples to the same array. We therefore developed a two-color labeling and detection scheme in which the reference is labeled with phycoerythrin (red), and the target with fluorescein (green) (13). By processing the reference and target together, experimental variability during the fragmentation, hybridization, washing, and detection steps is minimized or eliminated. In addition, during cohybridization of the reference and target, competition for binding sites results in a slight improvement in mismatch discrimination. Array hybridization is highly reproducible, and comparative analysis of data obtained from separate but identically synthesized arrays is also effective.

The two-color approach was tested by analyzing a 2.5-kb region of mtDNA that spans the tRNA<sup>Glu</sup>, cytochrome b, tRNA<sup>Thr</sup>, tRNA<sup>Pro</sup>, control region, and tRNA<sup>Phe</sup> DNA sequences (14). A  $P^{20,9}$  array (20-nucleotide oligomer probes varied at position 9 from the 3' end) was designed to match the mt1 target (that is,  $P^0$  sequence = mt1). The mt1 reference (red) and a polymorphic target sample (green) were pooled and hybridized simultaneously to the array. Differences between the target and reference sequences were identified by comparing the scaled red and green  $P^0$  hybridization intensities (15). The marked decrease in target hybridization intensity, over a span of  $\sim 20$  nucleotides, is shown for a single-base polymorphism at position 16,223 (Fig. 2A). The footprint is enlarged when two polymorphisms occur in close proximity (within  $\sim 20$  nucleotides) (Fig. 2B). When polymorphisms are clustered, the size of the footprint depends on the number of polymorphisms and their separation (Fig. 2C).



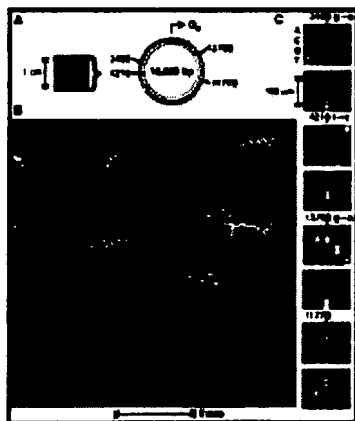
**Fig. 2.** Detection of base differences in a 2.5-kb region by comparison of scaled  $P^0$  hybridization intensity patterns between a sample (green) and a reference (red) sequence. (A) Comparison of sequence ief007 to mt1. In the region shown, there is a single-base difference between the two sequences, located at position 16,223 (C in mt1, T in ief007). This results in a "footprint" spanning  $\sim 20$  positions, 11 to the left and 8 to the right of position 16,223, in which the ief007  $P^0$  intensities are decreased by a factor of more than 10 on average relative to the mt1 intensities. The predicted footprint location is indicated by the gray bar, and the location of the

polymorphism is shown by a vertical black line within the bar. The size of a footprint changes with probe length, and its relative position with substitution position (not shown). (B) Comparison of sequence ha001 to mt1. The ha001 target has four polymorphisms relative to mt1. The  $P^0$  intensity pattern clearly shows two regions of difference between the targets. Each region contains two or more differences, because in both cases the footprints are longer than 20 positions and therefore are too extensive to be explained by a single-base difference. The effect of competition can be seen by comparing the mt1 intensities in the ief007 and ha001 experiments: The relative intensities of mt1 are greater in (B) where ha001 contains  $P^0$  mismatches but ief007 does not. (C) The ha004 sequence has multiple differences to mt1, resulting in a complex pattern extending over most of the region shown. Thus, differences are clearly detected. Because hybridization intensities are extremely sequence-dependent, each of the mitochondrial sequences can also be identified simply by its hybridization pattern. [View Larger Version of this Image (29K GIF file)]

To read polymorphisms accurately, we developed an algorithm that addresses the issue of multiple mismatches. The algorithm performs base identification but also flags regions of ambiguity caused by multiple mismatches. These regions are easily identified by the presence of a large footprint (Fig. 2, B and C) or by two or more bases identified as differing from  $P^0$  within the span of a single probe. Discrepancies between base identifications and footprint patterns are also flagged for further analysis (for example, a  $P^0$  footprint in which no polymorphism is identified; such a pattern is typical of a deletion). Thus, base identifications are valid only for unflagged regions. In flagged regions, the presence of sequence differences is detected, but no attempt is made to identify the sequence without further analysis.

Sequence analysis was carried out on the 2.5-kb target from 12 samples. A total of 30,582 bp containing 180 substitutions relative to mt1 were analyzed. Ninety-eight percent of the sequence was unambiguously assigned by a Bayesian base identification algorithm (16). Of this 98%, which contained both wild-type sequence and a high proportion of single-base footprints such as the example shown in Fig. 2A, 29,878 out of 29,879 bp were identified correctly (17). The remaining 2% of the sequence, which contained the multiple substitution footprints (such as those shown in Fig. 2, B and C), was flagged for further analysis. Of the 649 bp composing this 2%, 643 bp were located in or immediately adjacent to footprints (18). In all, 179 out of the 180 polymorphisms were unambiguously detected, 126 out of 127 were identified correctly in the unflagged regions, and 53 polymorphisms occurring in the flagged regions were detected as footprints. There were no unflagged false-positive base identifications, and only one false-positive footprint. These figures can be considered to be "worst case" estimates for the type of array and target used. The P<sup>0</sup> sequence represents a Caucasian haplotype, and our sample set included eight African samples having a large number of clustered differences to P<sup>0</sup>. Furthermore, the variation in the hypervariable part of the control region is much higher than for the rest of the mitochondrial genome and for nuclear genes in general (Fig. 2 shows comparisons to African samples in this region).

The determination of a complete human mitochondrial DNA sequence more than 15 years ago has had a tremendous influence on studies of human origins and evolution and the role of mutations in degenerative diseases (8, 10, 19). Because of the cost and difficulty of conventional sequence analysis, most subsequent sequencing studies have focused only on two small hypervariable regions totaling ~600 bp (9). However, access to the entire genome is required for a full understanding of the governing genetics. We therefore designed a P<sup>25,13</sup> tiling array for the mitochondrial genome. The array contains a total of 136,528 synthesis cells, each ~35  $\mu$ m by 35  $\mu$ m in size (Fig. 3). In addition to a 4L tiling across the genome, the array contains a set of probes representing a single-base deletion at every position across the genome and sets of probes designed to match a range of specific mtDNA haplotypes. Using long-range polymerase chain reaction, we amplified the 16.6-kb mtDNA directly from genomic DNA samples (20). Labeled RNA targets were prepared by in vitro transcription and hybridized to the array. Genomic hybridization patterns were imaged in less than 10 min by a high-resolution confocal scanner (21).



**Fig. 3.** Human mitochondrial genome on a chip. (A) An image of the array hybridized to 16.6 kb of mitochondrial target RNA (L strand). The 16,569-bp map of the genome is shown, and the H strand origin of replication (O<sub>H</sub>), located in the control region, is indicated. (B) A portion of the hybridization pattern magnified. In each column there are five probes: A, C, G, T, and  $\Delta$ , from top to bottom. The  $\Delta$  probe has a single-base deletion instead of a substitution and hence is 24 instead of 25 bases in length. The scale is indicated by the bar beneath the image. Although there is considerable sequence-dependent intensity variation, most of the array can be read directly. The image was collected at a resolution of ~100 pixels per probe cell. (C) The ability of the array to detect and read single-base differences in a 16.6-kb sample is illustrated.

Two different target sequences were hybridized in parallel to different chips. The hybridization patterns are compared for four different positions in the sequence. Only the P<sup>25,13</sup> probes are shown. The top panel of each pair shows the hybridization of the mt3 target, which matches the chip P<sup>0</sup> sequence at these positions. The lower panel shows the pattern generated by a sample from a patient with Leber's hereditary optic neuropathy (LHON). Three known pathogenic mutations, LHON3460, LHON4216, and LHON13708, are clearly detected. For comparison, the fourth panel in the set shows a region around position 11,778 that is identical in both samples. [View Larger Version of this Image (117K GIF file)]

The hybridization pattern of a 16.6-kb target to the mitochondrial genome chip is shown in Fig. 3. Although there are some regions of low intensity, most of the 25-nucleotide oligomer array hybridized efficiently: Simply by identifying the highest intensity in each column of four substitution probes, 99.0%

of the mt3 sequence could be read correctly ( $P^0$  sequence = mt3). The array was used to successfully detect three disease-causing mutations in a mtDNA sample from a patient with Leber's hereditary optic neuropathy (22, 23) (Fig. 3C). In addition, we detected a total of seven errors and new polymorphisms from previously unsequenced regions.

We then hybridized 10 genomes from African individuals to the array and unambiguously identified 505 polymorphisms. These were polymorphisms that could be clearly read and for which a confirmatory footprint was detected automatically. For the 10 samples, the 2.5-kb cytochrome b and control region sequences were known (17). No false positives were detected in the ~25 kb of sequence checked in this way. Additional clustered polymorphisms were detected by the presence of footprints but not read directly. A detailed analysis of the polymorphisms in these genomes, and others, will be presented elsewhere.

The throughput of a conventional gel-based sequencer, with an average read length of 400 nucleotides and 48 lanes that is run twice a day, might be two mitochondrial genomes a day at best. In contrast, the throughput of the nonoptimized system we describe is five chips per hour. Thus, 50 genomes can be read by hybridization in the time it takes to read two genomes conventionally. Furthermore, there are significant reductions in sample preparation requirements because the entire genome is labeled in a single reaction, so the cost is similar to that for a single sequencing reaction. Also, sequence reading at the level of data analysis is automated: The sequences can be read in a matter of minutes. No analytical separations or gel preparation is needed, which contributes to the speed of the experiment. Although the inability to read all possible sequences is a weakness of the 4L tiled array, it is not a major limitation, because in practice the small number of ambiguities can be checked by targeted conventional sequencing. In particular, highly repetitive sequences, such as long runs of a single base, are presently best analyzed with conventional technology. Finally, a clear advantage to the approach we describe is that it is highly scalable. The cost, effort, and time required to analyze the entire 16.6-kb mtDNA in a single experiment is virtually identical to that required to read 2.5 kb. This provides a clear path to further orders-of-magnitude improvements in efficiency.

High-density oligonucleotide arrays provide the foundation for a powerful genetic analysis technology. The method can be used to characterize the spectrum of sequence variation in a population and can be applied to the analysis of many genes in parallel. In the case of human mtDNA, we simultaneously analyzed the control region, 13 protein coding genes, 22 tRNA genes, and 2 ribosomal RNA genes. The methods described here can be applied to other research areas in molecular genetics; for example, the ability to identify and sequence polymorphisms provides a basis for genetic mapping. The specificity of oligonucleotide hybridization and the scalability of the method suggests the possibility of a dedicated array that could be used to generate a high-resolution genetic map of an entire genome in a single experiment. Likewise, the concepts and techniques described here have been used to develop approaches for mRNA identification and the large-scale, parallel measurement of expression levels (24). Thus, the sequence of a gene, its spectrum of change in the population, its chromosomal location, and its dynamics of expression (all essential to a full understanding of function) can be determined with high-density probe arrays. The challenge now is to synthesize and read probe arrays at even higher density. For example, a 2 cm by 2 cm array, synthesized with probes occupying 1- $\mu$ m synthesis sites in a 4L tiling, could query the entire coding content of the human genome, estimated at 100,000 genes.

---

## REFERENCES AND NOTES

1. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977) [Medline].
2. A. M. Maxam and W. Gilbert, *ibid.*, p. 560.
3. J. D. Watson and F. H. C. Crick, *Nature* **171**, 737 (1953).
4. W. Bains and G. C. Smith, *J. Theor. Biol.* **135**, 303 (1988) [Medline]; Y. P. Lysov *et al.*, *Dokl. Akad. Nauk. SSSR* **303**, 1508 (1988); R. Drmanac, I. Labat, I. Brukner, R. Crkvenjakov, *Genomics* **4**, 114 (1989) [Medline]; E. Southern, U. Maskos, R. Elder, *ibid.* **13**, 1008 (1992) [Medline]; see also R. B. Wallace *et al.*, *Nucleic Acids Res.* **6**, 3543 (1979) [Medline].

5. S. P. A. Fodor *et al.*, *Science* **251**, 767 (1991).
6. A. C. Pease *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5022 (1994) [Medline].
7. In the present format, we can routinely achieve a density of 409,600 synthesis sites in a 1.28 cm by 1.28 cm array. Each 20  $\mu$ m by 20  $\mu$ m site contains  $\sim 4 \times 10^6$  functional copies of a specific probe, which corresponds to a mean distance of about 100 Å between probes (M. O. Trulson, D. Stern, R. P. Rava, unpublished results).
8. S. Anderson *et al.*, *Nature* **290**, 457 (1981) [Medline].
9. The control region of mtDNA is characterized by high amounts of sequence polymorphism concentrated in two hypervariable regions [B. D. Greenberg, J. E. Newbold, A. Sugino, *Gene* **21**, 33 (1983) [Medline]; C. F. Aquardo and B. D. Greenberg, *Genetics* **103**, 287 (1983)].
10. R. L. Cann, W. M. Brown, A. C. Wilson, *Genetics* **106**, 479 (1984) [Medline].
11. The mt1 and mt2 sequences were cloned from amplified genomic DNA extracted from hair roots [P. Gill, A. J. Jeffreys, D. J. Werrett, *Nature* **318**, 577 (1985) [Medline]; R. K. Saiki *et al.*, *Science* **239**, 487 (1988) [Medline]]. The clones were sequenced conventionally (1). Cloning was performed only to provide a set of pure reference samples of known sequence. For templates for fluorescent labeling, DNA was reamplified from the clones with primers bearing bacteriophage T3 and T7 RNA polymerase promoter sequences (bold; mtDNA sequences uppercase): L15935-T3, 5'-ctcggaattaaccctcactaaaggAAACCTTTTCCAAGGA and H667-T7, 5'-taatacgactcactataggagAGGCTAGGACCAAACCTATT.
12. Labeled RNAs from the two complementary mtDNA strands [designated L and H (8)] were transcribed in separate reactions from a promoter-tagged polymerase chain reaction (PCR) product. Each 10- $\mu$ l reaction contained 1.5 mM each of the triphosphate nucleotides ATP, CTP, GTP, and UTP; 0.24 mM fluorescein-12-CTP (Du Pont); 0.24 mM fluorescein-12-UTP (Boehringer Mannheim);  $\sim 1$  to 5 nM (1.5  $\mu$ l) crude unpurified 1.3-kb PCR product; and T3 or T7 RNA polymerase (1 U/ $\mu$ l) (Promega) in a reaction buffer supplied with the enzyme. The reaction was carried out at 37°C for 1 to 2 hours. RNA was fragmented to an average size of <100 nucleotides by adjusting the solution to 30 mM MgCl<sub>2</sub>, by the addition of 1 M MgCl<sub>2</sub>, and heating at 94°C for 40 min. Fragmentation improved the uniformity and specificity of hybridization (M. Chee *et al.*, data not shown). The extent of fragmentation is dependent on the magnesium ion concentration [J. W. Huff, K. S. Sastry, M. P. Gordon, W. E. C. Wacker, *Biochemistry* **3**, 501 (1964); J. J. Butzow and G. L. Eichorn, *Biopolymers* **3**, 95 (1965) [Medline]]. Good hybridization results have been obtained with both DNA and RNA targets prepared with a variety of labeling schemes, including incorporation of fluorescent and biotinylated deoxynucleoside triphosphates by DNA polymerases, incorporation of dye-labeled primers during PCR, ligation of labeled oligonucleotides to fragmented RNA, and direct labeling by photo-cross-linking a psoralen derivative of biotin directly to fragmented nucleic acids (L. Wodicka, personal communication).
13. For two-color detection experiments, the reference and unknown samples were labeled with biotin and fluorescein, respectively, in separate transcription reactions. Reactions were carried out as described (12) except that each contained 1.25 mM of ATP, CTP, GTP, and UTP and 0.5 mM fluorescein-12-UTP or 0.25 mM biotin-16-UTP (Boehringer Mannheim). The two reactions were mixed in the ratio 1:5 (v/v) biotin:fluorescein and fragmented (12). Targets were diluted to a final concentration of  $\sim 100$  to 1000 pM in 3M TMACl [W. B. Melchior Jr. and P. H. von Hippel, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 298 (1973) [Medline]], 10 mM tris-HCl, pH 8.0, 1 mM EDTA, 0.005% Triton X-100, and 0.2 nM control oligonucleotide labeled at the 5' end with fluorescein (5'-CTGAACGGTAGCATCTTGAC). Samples were denatured at 95°C for 5 min, chilled on ice for 5 min, and equilibrated to 37°C. A volume of 180  $\mu$ l of hybridization solution was then added to the flow cell [R. Lipshutz *et al.*, *Biotechniques* **19**, 442 (1995) [Medline]] and the chip incubated at 37°C for 3 hours with rotation at 60 rpm. The chip was washed six times at room temperature with 6 $\times$  SSPE (0.9 M NaCl, 60 mM NaH<sub>2</sub>PO<sub>4</sub>, 6 mM EDTA, pH 7.4), 0.005% Triton X-100. Phycoerythrin-conjugated streptavidin (2  $\mu$ g/ml in 6 $\times$  SSPE, 0.005% Triton X-100) was added and incubation continued at room temperature for 5 min. The chip was washed again and scanned at a resolution of  $\sim 74$  pixels per probe cell. Two scans were collected: a fluorescein scan was obtained with a 515- to 545-nm band-pass filter, and a phycoerythrin scan with a 560-nm long-pass filter. Signals were separated to remove spectral overlap and average counts per cell determined.
14. Each 2.5-kb target sequence was PCR-amplified directly from genomic DNA with the primer pair

- L14675-T3 (5'-aattaaccctcactaaagggATTCTCGCACGGACTACAAC) and H667-T7 (11).
15. To scale the sample to the reference intensities, we constructed a histogram of the base 10 logarithm of the intensity ratios for each pair of probes. The histogram had a mesh size of 0.01 and was smoothed by replacing the value at each point with the average number of counts over a five-point window centered at that point. The highest value in the histogram was located, and the resulting intensity ratio was taken to be the most probable calibration coefficient.
  16. Base identification was accomplished with a Bayesian classification algorithm based on variable kernel density estimation. The likelihood of each identification associated with a set of hybridization intensity values was computed by comparing an unknown set of probes to a set of example cases for which the correct base identification was known. The resulting four likelihoods were then normalized so that they summed to 1. Data from both strands were combined by averaging the values. If the most likely base identification had an average normalized likelihood greater than 0.6, it was called, otherwise the base was called as an ambiguity. The example set was derived from two different samples, ib013 and ief005, which have a total of 35 substitutions relative to mt1, of which 19 are shared with the 12 samples analyzed and 16 are not. Identification performance was not sensitive to the choice of examples.
  17. To provide an independently determined reference sequence, each 2.5-kb PCR amplicon was sequenced on both strands by primer-directed fluorescent chain-terminator cycle sequencing with an ABI 373A DNA sequencer and assembled and manually edited with Sequencher 3.0. The analysis presented here assumes that the sequence amplified from genomic DNA is essentially clonal [R. J. Monnat and L. A. Loeb, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2895 (1985) [\[Medline\]](#)] and that its determination by gel-based methods is correct. A frequent length polymorphism at positions 303 to 309 was not detected by hybridization under the conditions used. It was excluded from analysis and is not part of the set of 180 polymorphisms discussed in the text. However, polymorphisms at this site have previously been differentiated by oligonucleotide hybridization [M. Stoneking, D. Hedgecock, R. G. Higuchi, L. Vigilant, H. A. Erlich, *Am. J. Hum. Genet.* **48**, 370 (1991) [\[Medline\]](#)].
  18. The  $P^0$  intensity footprints were detected in the following way: The reference and sample intensities were normalized (15), and  $R$ , the average of  $\log(P^0_{\text{reference}}/P^0_{\text{sample}})$  over a window of five positions, centered at the base of interest, was calculated for each position in the sequence. Footprints were detected as regions having at least five contiguous positions with a reference or sample intensity at least 50 counts above background and an  $R$  value in the top 10th percentile for the experiment. At 205 polymorphic sites, where the sample was mismatched to  $P^0$ , the mean  $R$  value was 1.01, with a standard deviation of 0.57. At 35,333 nonpolymorphic sites (that is, where both reference and sample had a perfect match to  $P^0$ ) the mean value was  $-0.05$ , with a standard deviation of 0.25.
  19. R. L. Cann, M. Stoneking, A. C. Wilson, *Nature* **325**, 31 (1987) [\[Medline\]](#); M. Zeviani *et al.*, *Am. J. Hum. Genet.* **47**, 904 (1990) [\[Medline\]](#); D. C. Wallace, *Annu. Rev. Biochem.* **61**, 1175 (1992) [\[Medline\]](#); S. Horai, K. Hayasaka, R. Kondo, K. Tsugane, N. Takahata, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 532 (1995) [\[Medline\]](#); T. Hutchin and G. Cortopassi, *ibid.*, p. 6892.
  20. Long-range PCR amplification was carried out on genomic DNA with Perkin-Elmer GeneAmp XL PCR reagents according to the manufacturer's protocol. Primers were L14836-T3 (5'-aattaaccctcactaaagggATGAAACTTCGGCTCACTCCTTGGCG) and RH1066-T7 (5'-taatacgactcactatagggATTTCATCATGCGGAGATGTTGGATGG), based on RH 1066 [S. Cheng, R. Higuchi, M. Stoneking, *Nature Genet.* **7**, 350 (1994) [\[Medline\]](#)]. Each 100- $\mu$ l reaction contained 0.2  $\mu$ M concentration of each primer and  $\sim$ 10 to 50 ng of total genomic DNA. Transcription reactions were carried out in 10  $\mu$ l with Ambion MAXIscript kit according to the manufacturer's protocol. The concentration of the 16.6-kb PCR template was  $\sim$ 2 nM, and the reaction contained Ambion 1 $\times$  biotin-14-CTP/NTP mix and 0.2 mM biotin-16-UTP. Incubation was at 37°C for 2 hours. Fragmentation and hybridization were as described (13), except that 3.5 M TMACl and the biotin-labeled oligonucleotide 5'-CTGAACGGTAGCATCTTGAC were used in the hybridization buffer, which also contained fragmented baker's yeast RNA (100  $\mu$ g/ml) (Sigma). Hybridization was carried out at 40°C for 4 hours.
  21. A custom telecentric objective lens with a numerical aperture of 0.25 focuses 5 mW of 488-nm argon laser light to a 3- $\mu$ m-diameter spot, which is scanned by a galvanometer mirror across a 14-mm field at 30 lines per second. Fluorescence collected by the objective is descanned by the

- galvanometer mirror, filtered by a dichroic beamsplitter (555 nm) and a band-pass filter (555 to 607 nm), focused onto a confocal pinhole, and detected by a photomultiplier. Photomultiplier output is digitized to 12 bits. A 4096 by 4096 pixel image is obtained in less than 3 min. Pixel size is 3.4  $\mu\text{m}$ . The data from four sequential scans were summed to improve the signal-to-noise ratio.
22. M. D. Brown, A. S. Voljavec, M. T. Lott, I. MacDonald, D. C. Wallace, *FASEB J.* 6, 2791 (1992)
  23. Mitochondrial DNA populations can contain more than one sequence type, in a condition known as heteroplasmy. The LHON mutations shown in Fig. 3C were characterized as being homoplasmic by conventional sequencing and restriction endonuclease digestion (M. Brown, personal communication). In controlled mixing experiments, we have shown that sequences present at the level of 10% can easily be detected by hybridization (M. Chee and R. Yang, unpublished results; N. Shen, personal communication). The sensitivity of detection is sequence dependent. Importantly, hybridization can be used to detect heterozygous nuclear DNA sequences (J. Hacia *et al.*, in preparation).
  24. D. J. Lockhart *et al.*, *Nature Biotech.*, in press.
  25. We thank M. Brown and D. Wallace for the gift of the LHON sample and R. Ward for the 10 African samples, M. Trulson for assistance in two-color hybridization, P. Fiekowsky for image analysis, and P. Berg and E. Lander for comments on the manuscript. R. Davis contributed to the initial concepts in oligonucleotide tiling. We especially thank L. Stryer for his incessant and persistent encouragement. Supported in part by Human Genome grant 5R01HG00813 from NIH (S.P.A.F.).

5 April 1996; accepted 26 July 1996

## This article has been cited by other articles:

- Inazuka, M., Wenz, H.-M., Sakabe, M., Tahira, T., Hayashi, K. (1997). A Streamlined Mutation Detection System: Multicolor Post-PCR Fluorescence Labeling and Single-Strand Conformational Polymorphism Analysis by Capillary Electrophoresis. *Genome Res.* 7: 1094-1103 [\[Abstract\]](#) [\[Full Text\]](#)
- Shaw, S. H., Carrasquillo, M. M., Kashuk, C., Puffenberger, E. G., Chakravarti, A. (1998). Allele Frequency Distributions in Pooled DNA Samples: Applications to Mapping Complex Disease Genes. *Genome Res.* 8: 111-123 [\[Abstract\]](#) [\[Full Text\]](#)
- Gingeras, T. R., Ghandour, G., Wang, E., Berno, A., Small, P. M., Drobniowski, F., Alland, D., Desmond, E., Holodniy, M., Drenkow, J. (1998). Simultaneous Genotyping and Species Identification Using Hybridization Pattern Recognition Analysis of Generic Mycobacterium DNA Arrays. *Genome Res.* 8: 435-448 [\[Abstract\]](#) [\[Full Text\]](#)
- Schimenti, J., Bucan, M. (1998). Functional Genomics in the Mouse: Phenotype-Based Mutagenesis Screens. *Genome Res.* 8: 698-710 [\[Abstract\]](#) [\[Full Text\]](#)
- Cuppens, H., Lin, W., Jaspers, M., Costes, B., Teng, H., Vankeerberghen, A., Jorissen, M., Droogmans, G., Reynaert, I., Goossens, M., Nilius, B., Cassiman, J.-J. (1998). Polyvariant Mutant Cystic Fibrosis Transmembrane Conductance Regulator Genes . The Polymorphic (TG)m Locus Explains the Partial Penetrance of the T5 Polymorphism as a Disease Mutation. *J. Clin. Invest.* 101: 487-496 [\[Abstract\]](#) [\[Full Text\]](#)
- Lashkari, D. A., McCusker, J. H., Davis, R. W. (1997). Whole genome analysis: Experimental access to all genome sequenced segments through larger-scale efficient oligonucleotide synthesis and PCR. *Proc. Natl. Acad. Sci. U. S. A.* 94: 8945-8947 [\[Abstract\]](#) [\[Full Text\]](#)
- Paces, V., Maltsev, N., Haselkorn, R., Fonstein, M. (1997). Sequence of a 189-kb segment of the chromosome of *Rhodobacter capsulatus* SB1003. *Proc. Natl. Acad. Sci. U. S. A.* 94: 9384-9388 [\[Abstract\]](#) [\[Full Text\]](#)
- Chen, X., Zehnbaauer, B., Gnirke, A., Kwok, P.-Y. (1997). Fluorescence energy transfer detection

- ▶ [PubMed Citation](#)
- ▶ [Related Articles in PubMed](#)
- ▶ [Download to Citation Manager](#)
- ▶ Search Medline for articles by:  
[Chee, M.](#) || [Fodor, S. P. A.](#)
- ▶ Alert me when:  
[New articles cite this article](#)

as a homogeneous DNA diagnostic method. *Proc. Natl. Acad. Sci. U. S. A.* 94: 10756-10761

[\[Abstract\]](#) [\[Full Text\]](#)

- Lee, M. (1998). Genome projects and gene pools: New germplasm for plant breeding?. *Proc. Natl. Acad. Sci. U. S. A.* 95: 2001-2004 [\[Abstract\]](#) [\[Full Text\]](#)
- Simpson, P. C., Roach, D., Woolley, A. T., Thorsen, T., Johnston, R., Sensabaugh, G. F., Mathies, R. A. (1998). High-throughput genetic analysis using microfabricated 96-sample capillary array electrophoresis microplates. *Proc. Natl. Acad. Sci. U. S. A.* 95: 2256-2261 [\[Abstract\]](#) [\[Full Text\]](#)
- Cho, R. J., Fromont-Racine, M., Wodicka, L., Feierbach, B., Stearns, T., Legrain, P., Lockhart, D. J., Davis, R. W. (1998). Parallel analysis of genetic selections using whole genome oligonucleotide arrays. *Proc. Natl. Acad. Sci. U. S. A.* 95: 3752-3757 [\[Abstract\]](#) [\[Full Text\]](#)
- Lander, E. S. (1996). The New Genomics: Global Views of Biology. *Science* 274: 536-539 [\[Full Text\]](#)
- Strauss, E. J., Falkow, S. (1997). Microbial Pathogenesis: Genomics and Beyond. *Science* 276: 707-712 [\[Abstract\]](#) [\[Full Text\]](#)
- Ouyang, Q., Kaplan, P. D., Liu, S., Libchaber, A. (1997). DNA Solution of the Maximal Clique Problem. *Science* 278: 446-449 [\[Abstract\]](#) [\[Full Text\]](#)
- DeRisi, J. L., Iyer, V. R., Brown, P. O. (1997). Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. *Science* 278: 680-686 [\[Abstract\]](#) [\[Full Text\]](#)
- Lin, V. S., Motesharei, K., Dancil, K. S., Sailor, M. J., Ghadiri, M. R. (1997). A Porous Silicon-Based Optical Interferometric Biosensor. *Science* 278: 840-843 [\[Abstract\]](#) [\[Full Text\]](#)
- Wang, D. G., Fan, J., Siao, C., Bero, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T. J., Lipshutz, R., Chee, M., Lander, E. S. (1998). Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome. *Science* 280: 1077-1082 [\[Abstract\]](#) [\[Full Text\]](#)

Volume 274, Number 5287 Issue of 25 Oct 1996, pp. 610 - 614

©1998 by The American Association for the Advancement of Science.

[SCIENCE NOW](#)

[HOME](#)

[HELP](#)

[SEARCH](#)

[ARCHIVE](#)

[CAREERS](#)

[NEXTWAVE](#)

[FEEDBACK](#)

[ORDER an article](#)

[MARKETPLACE](#)

[Copyright © 1998 by the American Association for the Advancement of Science.](#)



Carl Zeiss



Science  
MAGAZINE

HOME

HELP

SEARCH

ARCHIVE

SUBSCRIPTIONS

FEEDBACK

ORDER an article

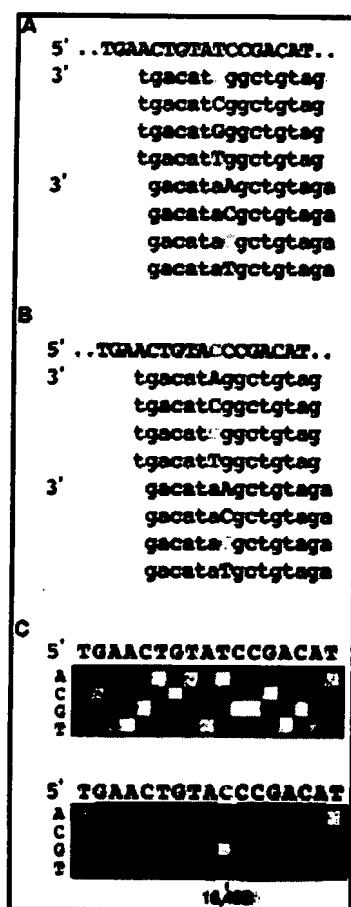
SIGN IN

SUSAN T ARNOLD || [Change Password](#) || [View/Change User Information](#) || [CiteTrack Personal Alerts](#) || [Subscription HELP](#) || [Sign Out](#)

[\[HELP with High Resolution Image Viewing\]](#) [\[Return to Article\]](#)

## Figure 1

[\[View Larger Version of this Image \(139K JPEG file\)\]](#)



**Figure 1.** (A) Design of a 4L tiled array. Each position in the target sequence (uppercase letters) is queried by a set of four probes on the chip (lowercase letters), identical except at a single position, termed the substitution position, which is either A, C, G, or T (blue indicates complementarity, red a mismatch). Two sets of probes are shown, querying adjacent positions in the target. (B) Effect of a change in the target sequence. The probes are the same as in (A), but the target now contains a single-base substitution (base C, shown in green). The probe set querying the changed base still has a perfect match (the G probe). However, probes in adjacent sets that overlap the altered target position now have either one or two mismatches (red) instead of zero or one, because they were designed to match the target shown in (A). (C) Hybridization to a 4L tiled array and detection of a base change in the target. The array shown was designed to the mt1 sequence. (Top) hybridization to mt1. The substitution used in each row of probes is indicated to the left of the image. The target sequence can be read 5' to 3'

from left to right as the complement of the substitution base with the brightest signal. With hybridization to mt2 (bottom), which differs from mt1 in this region by a T→C transition, the G probe at position 16,493 is now a perfect match, with the other three probes having single-base mismatches (A 5, C 3, G 37, T 4 counts). However, at flanking positions, the probes have either single- or double-base mismatches, because the mt2 transition now occurs away from the query position.

[\[Return to Article\]](#)

Volume 274, Number 5287 Issue of 25 Oct 1996, p 610

©1998 by The American Association for the Advancement of Science.

---

|                             |                          |                                  |                             |                         |                         |
|-----------------------------|--------------------------|----------------------------------|-----------------------------|-------------------------|-------------------------|
| <a href="#">SCIENCE NOW</a> | <a href="#">HOME</a>     | <a href="#">HELP</a>             | <a href="#">SEARCH</a>      | <a href="#">ARCHIVE</a> | <a href="#">CAREERS</a> |
| <a href="#">NEXTWAVE</a>    | <a href="#">FEEDBACK</a> | <a href="#">ORDER an article</a> | <a href="#">MARKETPLACE</a> |                         |                         |

[Copyright © 1998 by the American Association for the Advancement of Science.](#)

ESSAYS ON SCIENCE AND SOCIETY  
150 YEARS • 1848-1998

HP Chemical Analysis  
Trade Up Offer

Science  
MAGAZINE

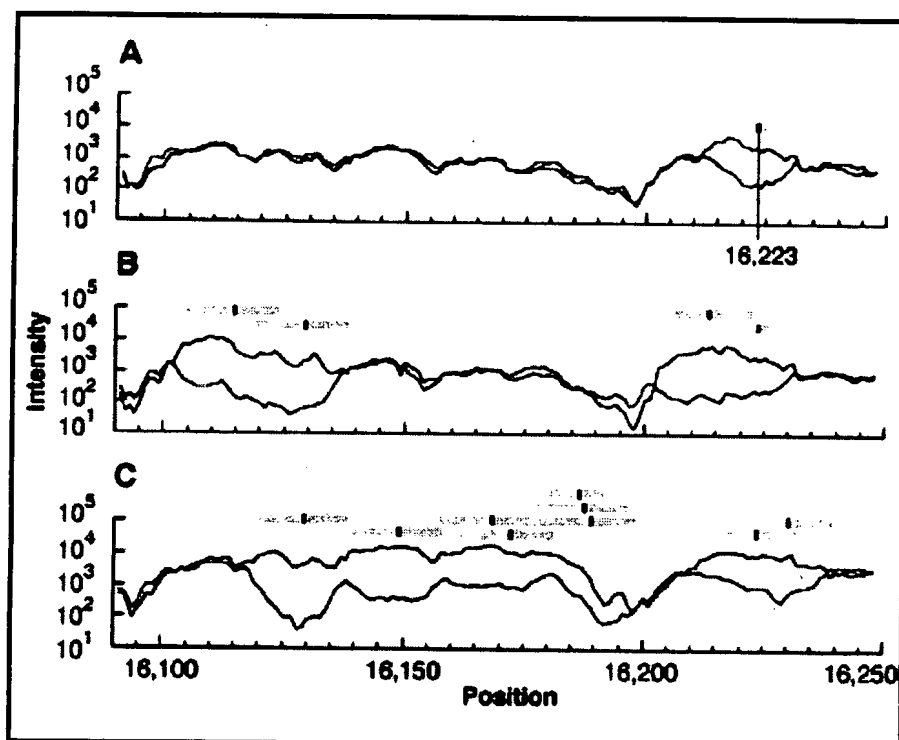
HOME HELP SEARCH ARCHIVE SUBSCRIPTIONS  
FEEDBACK ORDER an article SIGN IN

SUSAN T ARNOLD || [Change Password](#) || [View/Change User Information](#) || [CiteTrack Personal Alerts](#) || [Subscription HELP](#) || [Sign Out](#)

[\[HELP with High Resolution Image Viewing\]](#) [\[Return to Article\]](#)

## Figure 2

[\[View Larger Version of this Image \(141K JPEG file\)\]](#)



**Figure 2.** Detection of base differences in a 2.5-kb region by comparison of scaled  $P^0$  hybridization intensity patterns between a sample (green) and a reference (red) sequence. (A) Comparison of sequence ief007 to mt1. In the region shown, there is a single-base difference between the two sequences, located at position 16,223 (C in mt1, T in ief007). This results in a "footprint" spanning  $\sim 20$  positions, 11 to the left and 8 to the right of position 16,223, in which the ief007  $P^0$  intensities are decreased by a factor of more than 10 on average relative to the mt1 intensities. The predicted footprint location is indicated by the gray bar, and the location of the polymorphism is shown by a vertical black line within the bar. The size of a footprint changes with probe length, and its relative position with substitution position (not shown). (B) Comparison of sequence ha001 to mt1. The ha001 target has four polymorphisms relative to mt1. The  $P^0$  intensity pattern clearly shows two regions of difference between the targets. Each region contains two or more differences, because in both cases the footprints are longer than 20 positions and therefore are too extensive to be explained by a single-base difference. The effect of competition can be seen by comparing the mt1 intensities in the ief007 and ha001 experiments: The relative intensities of mt1 are greater in (B) where ha001 contains  $P^0$  mismatches but ief007 does not. (C) The ha004 sequence has multiple differences to mt1, resulting in a complex pattern extending over most of the

region shown. Thus, differences are clearly detected. Because hybridization intensities are extremely sequence-dependent, each of the mitochondrial sequences can also be identified simply by its hybridization pattern.

[\[Return to Article\]](#)

Volume 274, Number 5287 Issue of 25 Oct 1996, p 610

©1998 by The American Association for the Advancement of Science.

---

[SCIENCENOW](#)[HOME](#)[HELP](#)[SEARCH](#)[ARCHIVE](#)[CAREERS](#)[NEXTWAVE](#)[FEEDBACK](#)[ORDER an article](#)[MARKETPLACE](#)

[Copyright © 1998 by the American Association for the Advancement of Science.](#)

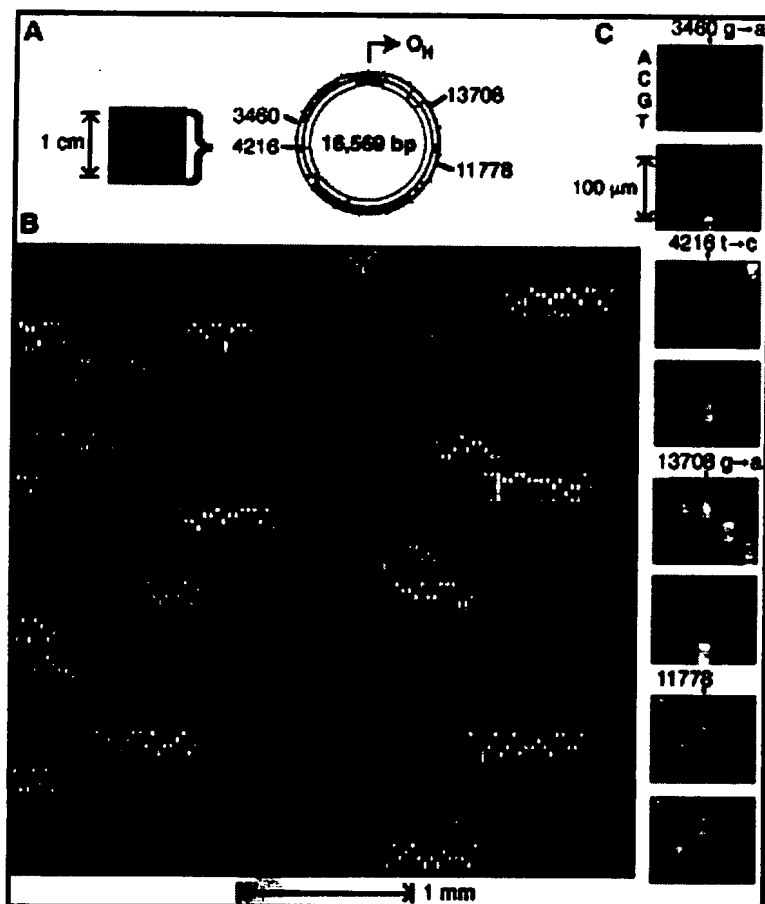
|             |                                      |                                     |                         |         |         |               |
|-------------|--------------------------------------|-------------------------------------|-------------------------|---------|---------|---------------|
| T<br>R<br>Y | Journal of<br>Neurophysiology Online | Clontech<br>Plasmid<br>Purification | We've Surf'd<br>For You |         |         |               |
|             | Science<br>MAGAZINE                  | HOME                                | HELP                    | SEARCH  | ARCHIVE | SUBSCRIPTIONS |
|             | FEEDBACK                             | ORDER an article                    |                         | SIGN IN |         |               |

SUSAN T ARNOLD || [Change Password](#) || [View/Change User Information](#) || [CiteTrack Personal Alerts](#) || [Subscription HELP](#) || [Sign Out](#)

[[HELP with High Resolution Image Viewing](#)] [[Return to Article](#)]

## Figure 3

[[View Larger Version of this Image \(273K JPEG file\)](#)]



**Figure 3. Human mitochondrial genome on a chip.** (A) An image of the array hybridized to 16.6 kb of mitochondrial target RNA (L strand). The 16,569-bp map of the genome is shown, and the H strand origin of replication ( $O_H$ ), located in the control region, is indicated. (B) A portion of the hybridization pattern magnified. In each column there are five probes: A, C, G, T, and  $\Delta$ , from top to bottom. The  $\Delta$  probe has a single-base deletion instead of a substitution and hence is 24 instead of 25 bases in length. The scale is indicated by the bar beneath the image. Although there is considerable sequence-dependent intensity variation, most of the array can be read directly. The image was collected at a resolution of  $\sim 100$  pixels per probe cell. (C) The ability of the array to detect and read single-base differences in a 16.6-kb sample is illustrated. Two different target sequences were hybridized in parallel to different chips. The hybridization patterns are compared for four different positions in the sequence. Only the

P<sup>25,13</sup> probes are shown. The top panel of each pair shows the hybridization of the mt3 target, which matches the chip P<sup>0</sup> sequence at these positions. The lower panel shows the pattern generated by a sample from a patient with Leber's hereditary optic neuropathy (LHON). Three known pathogenic mutations, LHON3460, LHON4216, and LHON13708, are clearly detected. For comparison, the fourth panel in the set shows a region around position 11,778 that is identical in both samples.

[\[Return to Article\]](#)

Volume 274, Number 5287 Issue of 25 Oct 1996, p 610

©1998 by The American Association for the Advancement of Science.

---

|                             |                          |                                  |                             |                         |                         |
|-----------------------------|--------------------------|----------------------------------|-----------------------------|-------------------------|-------------------------|
| <a href="#">SCIENCE NOW</a> | <a href="#">HOME</a>     | <a href="#">HELP</a>             | <a href="#">SEARCH</a>      | <a href="#">ARCHIVE</a> | <a href="#">CAREERS</a> |
| <a href="#">NEXTWAVE</a>    | <a href="#">FEEDBACK</a> | <a href="#">ORDER an article</a> | <a href="#">MARKETPLACE</a> |                         |                         |

Copyright © 1998 by the American Association for the Advancement of Science.